DOCUMENT RESUME

ED 043 680 TM 000 158

AUTHOR	Merwin, Jack C.
TITLE	Evaluation Designs and Instruments.
PUB DATE	Feb 69
NOTE	7p.; From symposium "The World of Evaluation Needs Re-Shaping" at the annual meeting of the American Educational Research Association, February 1969
EDRS PRICE	EDRS Price MF-$0.25 HC-$0.45
DESCRIPTORS	Curriculum Evaluation, *Decision Making, *Evaluation, Evaluation Methods, Evaluation Techniques, Measurement, Measurement Goals, Program Evaluation, *Testing

ABSTRACT

Available designs and instruments for evaluation are not adequate to meet the varied needs of evaluation. Therefore, evaluation needs to be reshaped in terms of 1) consideration of how and where currently available theories, designs, and instruments are proving useful, 2) identification of needs that cannot be met ith currently available constructs and tools, and 3) an attempt to identify guidelines for efforts to meet unfulfilled needs. On this context, the history of educational testing and evaluation is briefly reviewed and some interesting new ideas noted. The rather recent concepts of formative versus summative evaluation, of fidelity versus bandwidth of information and of group evaluation versus individual evaluation might be of help to reshape evaluation positively if the needs for evaluation can be examined within a framework of educational decision making. (CK)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED043680

Symposium: "The World of Evaluation Needs Re-shaping"*

Evaluation Designs and Instruments
Jack C. Merwin
University of Minnesota

I was happy to accept the chairman's invitation to participate in this symposium because I felt the title reflected many of my personal biases. Within the framework of our frustrations with available designs and instruments which do not meet many of our varied needs for evaluation, the term re-shaping implies to me, 1) consideration of where and how currently available theories, designs and instruments are proving useful, 2) identification of needs that cannot be met with currently available constructs and tools, and 3) an attempt to identify guidelines for efforts to meet unfulfilled needs.

In my brief comments this morning, I will attempt to put the dimensions of our current needs in a historical perspective. The most promising aspect of current frustration is the long overdue recognition that we can no longer live with the totally unrealistic idea that a small number of designs and a very limited variety of evaluative instruments can serve all of our needs for evaluation in education.

I view the following as encouraging signs of movement and trends toward the needed reshaping of the world of evaluation as it relates to evaluating individuals:

---

*American Educational Research Association Meeting, February 1969.

TM 000 158

1. Emphasis on measuring change, rather than status, many problems of which are brought out in a report of the Wisconsin Symposium, Problems in Measuring Change, edited by Chester Harris.
2. Explorations of the use of sequential procedures for gathering information, as opposed to across the board administration of instruments.
3. Experimentation with placement tests, "imbedded" items and proficiency tests as part of the learning process, such as that of the Oakleaf Project of Glaser and his associates.

On the latter of these points, it is interesting to note something similar from the past. Monroe's book of 1918, Measuring the Results of Teaching, carried a focus on mastery of skills related to very specific objectives.

Our evaluation efforts in recent decades have focused on evaluation of the individual and indeed there is further development and reshaping needed in this area. But there have been other needs for evaluation which have gone largely unheeded for some time. In his paper "Course Improvement Through Evaluation," Lee Cronbach describes the situation in this way:

> Many types of decisions are to be made, and many varieties of information are useful. It becomes immediately apparent that evaluation is a diversified activity and that no one set of principles will suffice for all situations. But measurement specialists have so concentrated upon one process--the preparation of pencil-and-paper achievement tests for assigning scores to individual pupils--that the principles pertinent to that process have somehow become enshrined as the principles of evaluation.

Much recent concern has not been with evaluation of individuals but with evaluation of programs; instruction, curriculum, methodology and so forth. Looking to the past first, we note that at the turn of the century there was a similar concern. Rice's classic study of the 1890's was aimed at a comparison of outcomes of different approaches to teaching the same subject. The 1916 NSSE Yearbook was entitled Standards and Tests for Measurement of the Efficiency of Schools and School Systems. That same year, Arnold produced a book entitled Measurement of Teaching Efficiency. In 1918, Monroe authored a book entitled Measuring the Results of Teaching, and the NSSE Yearbook for that year was The Measurement of Educational Products. It was with the background of design and instrumentation set forth in such books that the great expansion of achievement testing took place in the 1920's.

I believe Cronbach hit upon the basic reason for many of our frustrations today as we look to currently available designs and instruments for program evaluation. He wrote,

> At that time (1920), the content of any course was taken pretty much as established and beyond criticism save for small shifts of topical emphasis. At the administrator's discretion, standard tests covering the curriculum were given to assess the efficiency of the teacher or the school system. Such administrative testing fell into disfavor when used injudiciously and heavy handily in the 1920's and 1930's. Administrators and accrediting agencies fell back upon descriptive features of the school program in judging adequacy. Instead of collecting direct evidence of educational impact, they judged schools in terms of size of budget, student-staff ratio, square feet of laboratory space, and the number of advanced credits accumulated by the teacher.

In this article from the Teachers College Record in 1963, Cronbach's next sentence is "This tide, it appears, is about to turn." Today we are looking at the needs for evaluation designs and instruments from a somewhat different view than our predecessors of the 1920 era. We are concerned not only with effectiveness of teaching, but also the effectiveness of "innovations" in all aspects of education.

Since the 1930's testing has been almost exclusively designed for judgments about individuals. Summary figures across scores for individual have provided some information regarding program effectiveness. We have been all too long, however, in coming to the realization that this approach often is not only inefficient, but simply does not provide some of the information needed. Thus, whether we attribute it to requirements for evaluation written into federal legislation, new approaches to teaching, or numerous curriculum development projects, the pressure has mounted to produce a healthy concern about the need for reshaping evaluation methodology and instruments to implement that methodology.

Irritating as it is to face broadened evaluation needs and find that available tools will simply not do the job, several types of activity already started indicate movement in promising directions.

One such activity that I would cite is the proposed use of a decision-making framework as a basis for thinking about evaluation. Stufflebeam has been working specifically on educational decision making as a framework, and Cronbach and Gleser earlier had set forth a general background. Stake's paper, "The Countenance of Educational Evaluation" provided a refreshing new view. The attention being given to mastery testing by Glaser et al at Pittsburgh and Bloom in Chicago, along with

the work on "Universe-defined" tests by Osborne and by Hively have been interesting new developments. Cronbach's proposal for an unmatched design for collecting information from groups should be included in this list, as should the efforts toward unique designs and instrumentation that has been under development by the Committee on Assessing the Progress of Education. And, I should not end this listing without mentioning the AERA Committee on Curriculum Evaluation and the monograph series started by that Committee.

I also want to mention some concepts of relatively recent vintage that have not been in the focus of design and instrument development, but which may well help us in reshaping of the world of evaluation around design and instrumentation. One is the distinction between formative and summative evaluation set forth by Scriven. A second is the concept of fidelity versus bandwidth of information suggested by Cronbach and Gleser. A third is the general idea of group evaluation as opposed to individual evaluation. And, finally, I would propose that all of such concepts might most readily move us toward a positive reshaping of evaluation if our needs for evaluation can be examined within the framework of educational decision making.

REFERENCES

Arnold, F. The Measurement of Teaching Efficiency. New York: Lloyd Adams Noble, 1916.

Bloom, B.S. "Learning for Mastery." Evaluation Comment, U.C.L.A. - C.S.E.I.P., I, May, 1968.

Cronbach, L.J. "Course Improvement Through Evaluation." Teachers College Record, LXIV (May, 1963), 672-83.

Cronbach, L.J. and Gleser, G.C. Psychological Tests and Personnel Decisions. (Second Edition). University of Illinois Press, Urbana, Illinois, 1965.

Glaser, R. "Adapting the Elementary School Curriculum to Individual Performance." Proceedings of the 1967 Invitational Conference on Testing Problems, Princeton, New Jersey: Educational Testing Service, 1968, 3-36.

Hively, W. II, Patterson, H.L. and Page, S.H. "A Universe-Defined System of Arithmetic Achievement Tests." Journal of Educational Measurement, Winter 1968, 275.

Merwin, J.C. and Womer, F.B. "Evaluation in Assessing the Progress of Education to Provide Bases of Public Understanding and Public Policy." Educational Evaluation: New Roles, New Means, (Sixty-Eighth Yearbook of the National Society for the Study of Education, Part II), Distributed by the University of Chicago Press, Chicago, Illinois, 1969, 305-34.

Monroe, W.S. Measuring the Results of Teaching. Houghton Miffin Co., Boston, 1918.

Osburn, H.G. "Item Sampling for Achievement Testing." Educational and Psychological Measurement, XXVIII (1968), 95-104.

Rice, J.M. "The Futility of the Spelling Grind." Forum, XXIII, April, June 1897, 163-172, 409-419.

Scriven, M. "The Methodology of Evaluation." Perspectives of Curriculum Evaluation: American Educational Research Association Monograph Series on Curriculum Evaluation, Rand McNally and Co., Chicago, 1967, 39-83.

Stake, R.E. "The Countenance of Educational Evaluation." Teachers College Record, LXVII, April 1967, 523-40.

Standards and Tests for the Measurement of the Efficiency of Schools and School Systems. Fifteenth Yearbook of the National Society for the Study of Education, Part II, Distributed by the University of Chicago Press, Chicago, Illinois, 1916.

Stufflebeam, D.L. "A Depth Study of the Evaluation Requirement." Theory into Practice, V, June 1966, 121-33.

The Measurement of Educational Products. Seventeenth Yearbook of the National Society for the Study of Education, Part II, Public School Publishing Company, Bloomington, Illinois, 1918.